Two-component Response Regulators from Arabidopsis thaliana Contain a Putative DNA-binding Motif (MOLECULAR BIOLOGY AND INFORMATION-Molecular Biology)

AUTHOR(S):

Sakai, Hiroe; Aoyama, Takashi; Bono, Hidemasa; Oka, Atsuhiro

Two-component Response Regulators from *Arabidopsis thaliana* Contain a Putative DNA-binding Motif

Hiroe Sakai, Takashi Aoyama, Hidemasa Bono, and Atsuhiro Oka

An expression sequence tag database of higher plants was screened by *in silico* profile analysis for response regulators of the two-component regulatory system. Two closely related genes (*ARR1* and *ARR2*), corresponding to one of the extracted candidates, were isolated from *Arabidopsis thaliana*. The two genes were comparably expressed in all tissues, and at higher levels in the roots. The amino-terminal half of their translation products was highly conserved. This is where a phosphate receiver domain with the landmark aspartate residue and a putative DNA-binding domain were located. Their carboxyl-terminal halves, although less similar to each other, included glutamine-rich and proline-rich regions characteristic of the transcriptional activation domain of eukaryotes. This architecture resembles that of typical bacterial response regulators serving as transcription factors.

Keywords: *Arabidopsis* / Response regulator / Transcription factor / Two-component regulatory system

Extracellular stimuli received by living cells are processed through signal transduction pathways and result in orchestrated gene expression. A large number of intracellular signal transduction pathways have been studied with eukaryotic cells. Many plant proteins (or their genes) similar to the components involved in animal and fungus signal transduction pathways have been identified. Recently, the *Arabidopsis thaliana* genes *ETR1* and *CKI1*, which are implicated in ethylene and cytokinin responses, respectively, have been shown to code for proteins similar to the sensor of a two-component regulatory system, which is the ubiquitous signal transduction system in bacteria (1). This suggests that the bacterial type of signal transduction pathway, or a similar one, may possibly exist in plant cells. To elucidate an entire plant signal transduction pathway incorporating the two-component regulatory system, the functions executed by the plant response regulator components must be known. In this respect, an attempt has been made to clone plant response regulator genes, but no response regulator in which the signal receiver domain accompanies other known functional domains has been identified. We here show the presence of two *A. thaliana* response regulators with characteristics of transcription factors (2), like the majority of bacterial response regulators.

Using the profile method (3), the similarity score was calculated for each entry in the plant EST database. We extracted 21 EST sequences of *A. thaliana* and rice with high scores, and found that 11 of their central aspartate residues were accompanied by additional landmark resi-

MOLECULAR BIOLOGY AND INFORMATION — Molecular Biology —

Scope of research

Attempts have been made to elucidate structure-function relationships of genetic materials and various gene products. The major subjects are mechanisms involved in signal transduction and regulation of gene expression responsive to environmental stimuli, differentiation and development of plant organs, and plant-microbe interaction. As of December 1998, study is being concentrated on the roles of homeo domain proteins, MADS box proteins, and DDK response regulators of higher plants in developmental and signal transduction processes.

Prof
OKA, Atsuhiro
(D Sc)

Assoc Prof
AOYAMA, Takashi
(D Sc)

Instr
GOTO, Koji
(D Sc)

Students:
TSUKUDA, Mayumi (DC)
HOMMA, Takashi (DC)
OHGISHI, Maki (DC)
SAKAI, Hiroe (DC)
LIANG Yajie (DC)
YANO, Hiroyuki (DC)
UEDA, Yumi (MC)
OHASHI, Yohei (MC)

dues, the fore aspartate or rear lysine residues at distances suitable for the D-D-K signal receiver domain in the response regulator. These can be structurally classified into three groups I, II, and III. Since the group I candidates appeared to code for additional functional domains from their mRNA sizes (see below), we carried out cloning experiments for the corresponding genes with an *A. thaliana* genomic library. As a result, we isolated two closely related D-D-K receiver genes (*ARR1* and *ARR2*).

The starting and terminating sites of *ARR1* and *ARR2* transcription were estimated by 5'- and 3'-RACE analyses on a mixture of *A. thaliana* cDNA. The transcriptional unit was thus determined to be 3,093 bp long for *ARR1* and 3,508 bp long for *ARR2*. These sequence data were deposited in the DDBJ/EMBL/GenBank databases (AB016471 and AB016472). To determine the exon-intron organization of these two genes, PCR was done on an *A. thaliana* cDNA mixture with various sets of primers, and the resulting PCR fragments were sequenced. We found that *ARR1* and *ARR2* are split by four and five introns, respectively. The sizes of the mature mRNA for *ARR1* and *ARR2* were thus calculated to be 2,362 and 2,697 residues, respectively.

The *ARR1* and *ARR2* translation products consisted of 669 and 664 amino-acid residues, respectively. Both proteins had comparable domain architecture. The D-D-K signal receiver domain was located at the amino-terminal end of both (aa 37-150 for ARR1 and aa 28-141 for ARR2), and 95% of the amino-acid residues were identical. A potential nuclear localization signal followed the receiver domain of both ARR1 (aa 152-157) and ARR2 (aa 143-148). Another highly similar region (96% identity) was aa 236-299 of ARR1 and aa 215-278 of ARR2. This region showed strong resemblance to a segment found in various proteins with unknown functions. In addition, we noticed a weak but significant similarity of this region to the DNA-binding Myb oncoprotein, particularly to the potato Myb homolog MybSt1, and this region is called the Myb-like domain hereafter. In contrast to the strong resemblance of the amino-terminal half of the gene products, the carboxyl-terminal half (aa 300-669 of ARR1 and aa 279-664 of ARR2) was not as highly conserved, and only 52% of the amino-acid residues were identical when the appropriate gaps were incorporated. Although the amino-acid sequence of this domain itself showed no obvious similarity to other proteins, this domain, particularly the one in ARR1, was rich in glutamine. It also had higher than average amounts of proline, serine, and phenylalanine residues.

Northern blot hybridization analysis of the ARR1 and ARR2 transcripts with specific probes was carried out by using total RNA from roots, rosette leaves, cauline leaves, stems, flower buds/flowers, and siliques of *A. thaliana* that had been grown under the standard conditions. The results revealed that the expression patterns of the two genes are similar to each other, though the transcription level of *ARR1* is always slightly higher than that of *ARR2*, and that the two genes are transcribed in all tissues, and at higher levels in the roots. The mRNA sizes estimated from the Northern analysis were about 2.4 kb for *ARR1* and 2.8 kb for *ARR2*. These values are consistent with those calculated from the transcription unit and exon-intron organization of both genes.

To estimate the copy number of *ARR1* and *ARR2* on the *A. thaliana* chromosomes, Southern blot hybridization analysis was performed. Under high-stringency conditions, each probe of the entire coding-region produced a limited number of signal bands, which were expected from the genomic sequence. On the other hand, under low-stringency conditions several additional weak bands were detected. The bands with relatively higher signal intensities corresponded to the bands that were clearly visualized by the other probe under the high-stringency conditions. These results indicated that the *A. thaliana* genome contains one copy each of *ARR1* and *ARR2* and presumably a few additional cognate genes, constituting a small gene family.

To obtain information on localization of the ARR1 and ARR2 proteins in cells, their cDNAs were connected in-frame to GUS under the control of cauliflower mosaic virus 35S promoter, and then introduced into onion epidermal cells by the particle delivery system. Histological staining analysis revealed that either fusion protein is located in nuclei. Furthermore, when the D-D-K signal receiver domain was replaced by yeast GAL4 DNA binding domain, the resulting recombinant proteins conferred the ability to activate transcription of the reporter gene with the GAL4 target element in tobacco cells. These results strongly suggest that ARR1 and ARR2 actually work as a transcription factor in plant cells, as supposed from their architecture. This is the first report delineating plant response regulators in which the signal receiver domain accompanies other functional domains such as nuclear localization domain, transcriptional activation domain, and probable DNA-binding domain.

References

1. Oka A, *Tanpakushitsu-Kakusan-Kouso*, **40**, 1010-1021 (1995).
2. Sakai H, Aoyama T, Bono H, and Oka A, *Plant Cell Physiol.*, **39**, 1232-1239 (1998).
3. Dodd IB, Egan JB, *J. Mol. Biol.* , **194**, 557-564 (1987).